Abolghasemi, Mahdi; Hurley, Jason; Eshragh, Ali; Fahimnia, Behnam. “Demand forecasting in the presence of systematic events: cases in capturing sales promotions”. *International Journal of Production Economics* Vol. 230, Issue December 2020, no. 107892 (2020).

Available from: http://dx.doi.org/10.1016/j.ijpe.2020.107892

Demand Forecasting in the Presence of Systematic Events: Cases in Capturing Sales Promotions

Mahdi Abolghasemi[1], Jason Hurley[2], Ali Eshragh[1], Behnam Fahimnia[2]
[1]School of Mathematical and Physical Sciences, The University of Newcastle, NSW, Australia.
[2]Institute of Transport and Logistics Studies, The University of Sydney, NSW, Australia

Abstract

Reliable demand forecasts are critical for effective supply chain management. Several endogenous and exogenous variables can influence the dynamics of demand, and hence a single statistical model that only consists of historical sales data is often insufficient to produce accurate forecasts. In practice, the forecasts generated by baseline statistical models are often judgmentally adjusted by forecasters to incorporate factors and information that are not incorporated in the baseline models. There are however systematic events whose effect can be quantified and modeled to help minimize human intervention in adjusting the baseline forecasts. In this paper, we develop and test a novel regime-switching approach to quantify systematic information/events and objectively incorporate them into the baseline statistical model. Our simple yet practical and effective model can help limit forecast adjustments to only focus on the impact of less systematic events such as sudden climate change or dynamic market activities. The model is validated empirically using sales and promotional data from two Australian companies. The model is also benchmarked against commonly employed statistical and machine learning forecasting models. Discussions focus on thorough analysis of promotions impact and benchmarking results. We show that the proposed model can successfully improve forecast accuracy and avoid poor forecasts when compared to the current industry practice which heavily relies on human judgment to factor in all types of information/events. The proposed model also outperforms sophisticated machine learning methods by mitigating the generation of extremely poor forecasts that drastically differ from actual sales due to changes in demand states.

Keywords*:* Demand Forecasting; Systematic Events; Time Series Regression Models; Sales Promotions; Judgmental Forecasting; Supply Chain.

1 Introduction

Demand forecasts[1] are critical pieces of information in supply chain management because numerous decisions – such as sourcing, production planning, logistics, inventory management and retail decisions – heavily rely on forecasts. In particular, demand forecasting is a key ingredient in sales and operations planning (S&OP) which is responsible for continuous alignment between demand plans and supply plans (Fildes, Goodwin & Önkal, 2018). Therefore, improving the accuracy of product demand forecasts can directly result in better operational efficiency, customer satisfaction, and financial savings throughout the entire supply chain (Kremer, Siemsen & Thomas, 2015; Trapero, Kourentzes & Fildes, 2015).

Having historical demand information is beneficial for generating accurate forecasts, albeit is often not solely sufficient to forecast to a desired degree of precision (Hyndman & Athanasopoulos, 2014). This is because many statistical forecasting models premised on historical data lack the ability to explicitly capture the contextual information[2] and/or dynamically update as more recent information becomes available (Lawrence et al., 2006). Special events such as marketing campaigns, holidays and sales promotions are examples of valuable information often not incorporated into univariate statistical forecasting models. Particularly, retailer sales promotions have been shown to significantly influence consumer behavior and market demand (Trapero, Kourentzes & Fildes, 2015; Trapero et al., 2013). Since events such as promotions can lead to non-stationary time series, single static time series forecasting methods may not be the most suitable. Hence in practice, the output of such methods are merely used as baseline forecasts which are subject to judgmental adjustment by demand forecasters (Fildes et al., 2009).

Using expert judgment in complement to statistically analyzing large amounts of data has been shown to be beneficial for improving forecast accuracy (Alvarado-Valencia et al., 2017). Moreover, evidence suggests that the human input to forecasts can be improved by providing a systematic approach to structure the information utilized when imposing judgment to make adjustments (Franses & Legerstee, 2013). For instance, a structured-analogies method is shown to lead to more accurate

[1] It is important to note that demand and sales are not always equal (e.g., sales are less than the true demand when stockouts occur) (Perera et al. 2019; Tong, Feiler & Larrick, 2018). However, they are treated as equal in the analysis performed in this paper because the companies investigated achieve near flawless service levels of 99.5%, on average.

[2] *Contextual information* refers to non-time series information that is highly relevant to interpreting, explaining and anticipating time series behavior.

forecasts than when produced with unaided judgment (Green & Armstrong, 2007). Yet no optimal procedure exists for structuring the critical information that forecasters must consider and there is no definitive answer regarding how to most effectively integrate human judgment with forecasting models (Baecke, De Baets & Vanderheyden, 2017).

In this paper, we aim to contribute to this area by developing and validating a forecasting model that can capture the effects of quantifiable systematic events which would otherwise require judgmental adjustments. It is important to note that this model is suited for well-defined and well-structured systematic events and not well suited for dynamic information that must be incorporated on a case to case basis. More precisely, we develop a time series regression model that incorporates some of the most significant factors/information considered by forecasters when adjusting baseline forecasts. The model takes into account the dynamics of historical base sales and the additional influential factors. Our observations in the Fast-Moving Consumer Goods (FMCG) industry have revealed that retailer sales promotions are the main reason for forecast adjustments and also most challenging to forecast. This is also supported by the academic literature (e.g., Fildes & Goodwin, 2007; Fildes et al., 2009). Therefore, we utilize the proposed model to deal with sales promotions, a good exemplar of systematic events. The model is able to systematically define various states of demand uplift by analyzing historical sales data and different combinations of promotions. The obtained states can then be embedded into the model. Although we explore the application of our model to deal with sales promotions, the underlying algorithm could be adapted to consider other systematic events such as holidays and seasonality trends.

The sales data is obtained from two large FMCG companies in Australia. The proposed model is tested using 253 time series from the case companies. The performance of the proposed model is compared with the current industry cases and well-established statistical and machine leaning models. Even though both businesses operate within the food and beverage industry, their product and promotion characteristics substantially differ in terms of product perishability and usage, as well as promotion frequency and magnitude of demand uplift, which to some extent helps the generalizability of our model application and study findings. This model can potentially aid sales forecasters and demand planners by reducing the complexity of the forecasting task and ease the cognitive load associated with processing vast amounts of unmodeled information (Lawrence et al., 2006).

The remainder of this paper is structured as follows. Section 2 reviews the literature related to demand forecasting in a supply chain context, including the main quantitative and judgmental approaches as well as how they handle the impact of sales promotions. Section 3 describes the methodology employed in this study and explains the structure of the proposed time series regression model. Section 4 focuses on validating the model and methodology using two empirical case studies. Concluding remarks are presented in Section 5 including the study limitations and future research directions.

2 Related Literature

A myriad of forecasting methods and models have been developed with the common goal of improving accuracy. In the following subsections, we review some of the quantitative (statistical and analytical) and judgmental forecasting approaches that are particularly concerned with addressing the impact of sales promotions.

2.1 Quantitative forecasting methods

Quantitative forecasting methods has evolved with advancements in computational power, software and information system technologies such as enterprise resource planning systems, electronic data interchange, and point of sale scanning (Sanders & Manrodt, 2003a). Such advancements have enabled vast amounts of data/information to be easily collected, utilized in more sophisticated statistical models, and shared throughout the supply chain. The core of most quantitative approaches to forecasting is extrapolation (Fildes et al., 2008). Extrapolative methods use purely historical data to predict the future. Of the extrapolative methods, exponential smoothing is one of the classical forecasting techniques and is widely practiced in industry (Fildes, 1992; Hyndman & Koehler, 2006). Exponential smoothing is a statistical technique that averages (smooths) time series data, differing from a simple moving average in that it assigns a larger weight to recent observations, and exponentially decreases the weight of observations over time. Several different variations of exponential smoothing exist (e.g., simple, Holt, Pegels, Holt-Winters, and variants of these with damped trends), where each method is suited for different forecast horizons (i.e., short-term or long-range), seasonality types and trends in the time series (i.e., constant, additive or multiplicative) (Taylor, 2003). Furthermore, the Auto-Regressive Integrated Moving Average (ARIMA) model (also known as Box-Jenkins model) along with its numerous variants (e.g., SARIMA, ARIMAX, ARMA-GARCH, ARFIMA) are also widely utilized extrapolative methods as they can account for trends, seasonality, errors and non-stationary aspects of a time series (Nikolopoulos et al., 2011).

The second main statistical forecasting approach is causal and multivariate methods, which have been largely developed by the study of econometrics data (Fildes et al., 2008). These methods are forms of regression analysis and assume that there is a cause-and-effect relationship between the dependent variable (i.e., demand) and one or more independent variables (explanatory factors influencing the demand). Causal and multivariate methods are capable of addressing the issue of a non-stationary time series by considering different exogenous variables such as promotions, holidays and special events that can impact customer demand (Fildes et al., 2008; Trapero et al., 2013). While there are many variables and vast amounts of information that can be included in forecasts, it is prudent to keep models as parsimonious as possible while maintaining desired accuracy. This is because multivariate models with a high number of explanatory variables have large data requirements, in addition to being prone to multicollinearity and dimensionality problems (Trapero, Kourentzes & Fildes, 2015).

Despite the rapid evolution of technology and numerous advancements in statistical and analytical forecasting methods, large-scale empirical evidence from three seminal forecasting competitions (Makridakis et al., 1982; Makridakis et al., 1993; Makridakis & Hibon, 2000) consistently finds that "statistically sophisticated or complex methods do not necessarily produce more accurate forecasts than simpler ones" (Makridakis & Hibon, 2000, p. 452). This notion is also supported by Green and Armstrong (2015) who find that complexity of the forecasting method harms accuracy, and that simpler methods reduce the likelihood of errors as well as better aid the understanding of decision-makers. Furthermore, complex forecasting techniques are not frequently utilized in industry due to high costs, lack of internal expertise and resources, as well as other organizational barriers (Trapero, Kourentzes & Fildes, 2015). We aim to address this concern in our study by presenting a simple and practical statistical model which captures the impact of systematic events, and yet still allows the forecaster to intervene and judgmentally incorporate the impact of less quantifiable contextual information. Therefore, the model and approach presented in this paper can be used as a more complete baseline forecast which would require fewer judgmental adjustments, if any at all. A substantial amount of research has been conducted over the years pertaining to judgmental forecasting and the integration of human judgment into statistical models. We briefly review this literature in Section 2.2.

2.2 The human factor in forecasting

Academic literature has increasingly acknowledged the importance of human judgment in forecasting as they are highly connected (Lawrence et al., 2006). Despite the broadly acknowledged human factor in forecasting, much of the research on forecasting methods that emerged prior to the 1990's advised against the use of judgment in forecasting (e.g., Armstrong, 1986; Hogarth & Makridakis, 1981). However, recent research advocates that statistical methods and human judgment should be integrated so that complementary benefits can be realized to mitigate the inherent weaknesses of each approach (Alvarado-Valencia et al., 2017; Baecke, De Baets & Vanderheyden, 2017; Blattberg & Hoch, 1990; Fischer & Harvey, 1999; Franses, 2008; Marmier & Cheikhrouhou, 2010).

One important benefit of integrating statistical methods and human judgment relates to their capability to handle different types of information. Lawrence, O'Connor and Edmundson (2000) classify the information that is useful for forecasting into two classes: (1) historical data, and (2) contextual or domain knowledge. Historical data is simply the time series of historical product sales that has been recorded, and contextual knowledge being any other information relevant to interpreting, explaining and anticipating time series behavior. Examples of contextual information include: changes in promotional plans, competitor activities, market intelligence, sudden climate changes and dynamic influencers (e.g., political, media/press release, natural or manmade disasters).

Contextual information is the primary factor that leads to instances when judgment is superior to statistical models (Webby & O'Connor, 1996). In fact, the major value-add that comes from human input is because forecasters possess contextual information, intimate product knowledge and experience that statistical models do not (Edmundson, Lawrence & O'Connor, 1988). Statistical methods are well suited to handle vast amounts of historical data, but when the effects of discontinuities (often caused by contextual factors) cannot be estimated from historical data, statistical methods tend to produce forecasts with sub-optimal accuracy (Gardner, 1985). Human judgment can be utilized to overcome this issue and incorporate valuable contextual information by adjusting baseline statistical forecasts, albeit with caution as there are also human factors that can hinder judgment (e.g., personal or social biases, heuristics, cognitive limitations, and system neglect) (Goodwin, 2002; Kremer, Siemsen & Thomas, 2015; Lawrence et al., 2006).

The great efficiency and flexibility as well as demonstrated accuracy improvements that are realized through judgmental forecast adjustments (e.g., Fildes et al., 2009; Moritz, Siemsen & Kremer, 2014) have inevitably resulted in its widespread use in industry (Sanders & Manrodt, 2003b). Overall, the debate surrounding the practice of judgmental forecast adjustments has now evolved beyond merely whether they should be utilized or abandoned. The question is more how to appropriately use judgment to consistently improve the accuracy of forecasts. A common practical issue is that forecast adjustments are rarely performed systematically (Trapero et al., 2013), with experts often applying their knowledge and experience in an unaided and unstructured form (Green & Armstrong, 2007). This may lead to poor forecasting outcomes when compared to a structured approach. Ideally, a Forecast Support System (FSS) can be utilized to provide a baseline statistical forecast as well as structured guidance and feedback to effectively inform a forecaster (Fildes, Goodwin & Lawrence, 2006; Goodwin et al., 2011). The use of an FSS helps ease the cognitive burden on the human mind and consequently improve the accuracy of final forecasts (Adya & Lusk, 2016). The model developed and tested in this paper can act as the core of such an FSS. Adding additional features such as adaptive guidance to this base model can help develop a comprehensive and practical FSS.

2.3 Promotional effects

Sales promotions are common phenomena in contemporary retail operations. Evidence suggests that promotions are the leading cause for judgmental adjustments to statistical forecasts (Fildes & Goodwin, 2007; Goodwin, 2002). When a promotion occurs, a price discount is offered to customers for a specified time-period and a variety of additional actions are also taken to increase the prominence of a given product or service. The additional actions taken are associated with the *promotional mechanics*, which may include: type of promotion (e.g., single-buy, buy one get one free, multi-buy), display type (e.g., front of store, end of aisle), advertisement type (e.g., in-store, online, catalogue), and special features to coincide with holidays/events (e.g., Christmas oriented product labelling, free event-oriented gift with purchase).

There is normally an uplift in sales when promotions are offered. The uplift is often associated with purchasing acceleration, increased consumption and/or brand switching (Blattberg & Neslin, 1989). Moreover, consumers commonly stockpile products while they are on promotion (that is more the case for less perishable items) which often leads to lower sales in the following period(s). Different combinations of promotions result in different sales uplift, but the magnitude of the impact is associated with a high degree of uncertainty given the dynamic nature of consumer behavior.

Inevitably, such promotional effects complicate the forecasting process.

The impact of sales promotions on demand has been previously explored and various models have been proposed (see for example Ali et al. (2009); Nikolopoulos et al. (2015); Ramanathan (2012); Ramanathan and Muyldermans (2011); Trapero et al. (2013)), yet quantifying the impact of promotions still proves to be problematic for practicing forecasters and academic researchers alike. There are several reasons why human judgment has been used in promotional sales forecasting. First, univariate statistical methods (e.g., exponential smoothing) only consider historical data and therefore do not account for the effects of future sales promotions in forecasts, unless promotions and corresponding effects are very consistent over time (Trapero, Kourentzes & Fildes, 2015). Although such methods are well suited for semi-automatically generating forecasts for numerous products, subsequent judgmental adjustment to account for contextual information is required. Second, judgment is particularly useful when little or no historical data is available such as when a new product or promotional campaign is offered (Oliva & Watson, 2009; Seifert et al., 2015). Judgment could also be beneficial in situations where large spikes in demand occur (significant promotions), because univariate statistical models disperse the effects of large changes over the entire horizon (Sanders, 1992). This can result in inaccurate parameter estimation for promotional versus non-promotional periods.

Sophisticated causal methods have also been proposed to handle the task of promotional forecasting (Fildes et al., 2008). These models are usually based on multiple regression with exogenous variables corresponding to various types of promotions. As opposed to judgmental forecasting which is practical and require few resources, such methods are highly complex, have demanding data requirements, and are difficult to interpret in terms of distinguishing the impact of individual promotional variables (Blattberg & Neslin, 1989; Trapero, Kourentzes & Fildes, 2015). Nevertheless, sophisticated forecasting models that account for the effects of promotions have been developed (Huang, Fildes & Soopramanien, 2014; Kourentzes & Petropoulos, 2016), in addition to promotional FSSs such as 'SCAN*PRO' (Van Heerde, Leeflang & Wittink, 2002), 'PromoCastTM' (Cooper et al., 1999), and 'CHAN4CAST' (Divakar, Ratchford & Shankar, 2005). There have also been attempts to innovate promotional modeling techniques by utilizing structural equation models (Ramanathan & Muyldermans, 2010) and dynamic regression involving principal component analysis and transfer functions (Trapero, Kourentzes & Fildes, 2015). Despite all those efforts, evidence indicates that lack of resources, expertise, and high costs hinder the widespread implementation of such methods and support systems in practice (Hughes, 2001).

We aim to tackle this issue by introducing an easy-to-implement and practical model that can be used to incorporate the impact of systematic promotions into the statistical models. Promotions are a good example of systematic events, the impact of which could be quantified and incorporated into the statistical models to provide the forecaster with a more solid and accurate forecast. This enables a forecaster to only focus on incorporating dynamic information and less systematic events whose level of impact requires expert opinion and market intelligence. Our aim in this paper is to develop and validate a new model that is 'simple' and yet 'practical and robust' to produce baseline statistical forecasts which may then only require minor judgmental adjustments. An interesting feature of the proposed model is its ability to avoid large errors and poor performance when forecasting sales of substantially different scales (e.g., promotional and non-promotional sales). The methodology utilized to do so is discussed next.

3 Methodology

In the previous section we described some of the key demand forecasting approaches and how sales promotions can dramatically alter the behavior of consumer demand. We have realized that similar scenarios where time series behavior is subject to constant changes have been studied in a macroeconomic context (e.g., Giordani & Kohn, 2008; Giordani, Kohn & van Dijk, 2007; Hamilton, 1989; Hamilton, 1990; Kim, Piger & Startz, 2008) and some of techniques and concepts utilized to tackle those situations could be applicable to a demand forecasting context. In particular, Hamilton (1989) developed a novel approach, the so-called Markov switching model, to more accurately capture and predict changes in the regime or state[3] of non-stationary time series. Although, Hamilton initially applied his model to the United States gross national product data, his approach has provided the foundation to forecast the future values of any time series that exhibits a regime-switching behavior.

We posit that in a product demand forecasting context, sales promotions can cause the time series to

[3] The terms 'regime' and 'state' are used interchangeably throughout this paper and are defined as "episodes across which the dynamic behavior of the series is markedly different" (Hamilton 1989, p. 358).

abruptly enter different states. Therefore, the concept of Hamilton's model can be applied to define demand states by which we can structure systematic promotional information and embed it in a statistical forecasting model. Motivated by Hamilton's regime switching idea, we develop a time series regression model by exploiting the time series data and information related to systematic events to capture sales dynamics in all periods and forecast future demand. To the best of our knowledge, this is the first attempt in demand forecasting literature to use a regime-switching approach to quantify the impact of systematic events.

Initially, a list of potential systematic events along with their associated levels of impact on product sales can be obtained from a company's forecasting experts. Given that these are often the same events that forecasters are already making routing judgmental adjustments to account for, it is reasonable to assume forecasters can generate a list of such events. Once identified, the most significant systematic events and levels can be verified through statistical analysis (e.g., Analysis of Variance (ANOVA)). The most significant systematic event for both the companies investigated in this paper is sales promotions. Therefore, all data pertaining to the various promotions ran for each product was gathered; including the *promotion type*, *display type* and *advertisement type*. This included information regarding the associated levels of impact (e.g., the 'promotion type' can take two possible levels: major and minor). Regression models were then constructed, validating the significance of each promotional variable with F-tests (ANOVA) by adding one at a time and comparing each new model with the previous one. If the variable was not significantly improving the fit of the model, it was removed.

Next, combinations of the levels of the most significant systematic events can be constructed, where each combination is called a *state*. This is a crucial element of our model and an important contribution of this paper because the procedure to establish various Demand Uplift States (DUS) can be done systematically. Sales uplifts in promotional periods are computed by subtracting baseline forecasts from the actual realized sales value in each epoch. In this research, we use historical baseline forecasts and the corresponding realized sales figures provided by two case companies. Both case companies generate their baseline forecasts using simple exponential smoothing. However, baseline forecasts can be estimated with numerous models such as those discussed in Section 2.1 (e.g., ARIMA, causal and multivariate methods). The DUS algorithm is summarized as follows:

Step 0. Input demand time series and a list of potential systematic events along with their possible levels.

Step 1. Find the most significant systematic events by running ANOVA over the data.

Step 2. Construct all possible combinations of the levels of significant systematic events and label them from *1* to *k*, where *k* is the total number of combinations.

Step 3. For i=1 to *k* do

Compute the average demand uplift for the i^{th} combination.

end for

Step 4. Put each combination with distinct uplifts in demand in one state.

return Demand uplift states labelled *1* to *m*, where *m* is the total number of states constructed in Step 4.

Let us consider sales promotions in retail industry as a case example for a significant systematic event. The common practice in the retail industry is for the retailers and suppliers to negotiate and set promotional plans well in advance of their occurrence. For instance, the case companies in this paper lock in their promotional calendars for the following calendar year. In addition to timing of each promotion, the specific promotion details (also referred to as promotional mechanics) are decided. However, promotional plans may be altered during the year for a variety of reasons. Examples include extending the well-performing promotions, adding extra promotions to induce sales for stock that is near expiration, and changing dates of promotions due to inclement weather. But final changes are normally locked at least four weeks prior to the promotion commencement. Therefore, given that promotional plans are finalized prior to preparing a forecast, the DUS algorithm can be effectively utilized to determine future states. If any change in promotional plans is realized, the algorithm can be easily re-run to accommodate abrupt variations. If a new significant level of the combination of promotions is offered, based on the potential impact on consumer behavior, forecasters can either assign it to one of the current demand uplift states (Step 4 of the DUS algorithm), or define a new state depending on the magnitude of its effect.

Following the DUS algorithm, we introduce our new time series regression model – the Forecasting Systematic Events (FSE) model – formulated in Equation 1.

$$X_t = \alpha_0 + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \sum_{j=1}^{m} \beta_j S_{jt} + \varepsilon_t \quad (1)$$

In this equation, X_t is the demand at time t, S_{jt} is the demand uplift state variable at time t (taking value of one if demand uplift is in state j at time t and zero otherwise), ε_t represents a Gaussian White

Noise process, α_0 is constant, ϕ_i , θ_i and β_j are unknown parameters that will be estimated by the time series data and the DUS algorithm output, p is the number of past sales values regressed in the model, q is the order of moving average, and m is the total number of sales uplift states prescribed by the DUS algorithm.

The FSE model has two components: (i) the first two terms on the right-hand side, which form an ARIMA model of order p and q to capture the underlying time series (i.e., the time series in the absence of systematic events), and (ii) regression over the demand uplift state variable S_{jt} to model the effect of systematic events. When there is no event at time period t, all state variables in that time period are equal to zero and the FSE model is simplified to an ARIMA model of order p and q. The FSE model assumes that the underlying time series is stationary. Consequently, if there exists a trend in the mean of the underlying time series, the sales variable X_t can be replaced with a sales variable differenced in an appropriate lag to convert it to a stationary time series. Importantly, the FSE model differs from a conventional ARIMA model with exogenous regressors (ARIMAX). Although the FSE model uses ARIMA to estimate the baseline model, it utilizes the DUS algorithm to determine and embed the regressors as opposed to ARIMAX that concurrently fits an ARIMA model to the entire dataset and uses regressors to account for changes in the series' dynamics. To demonstrate the application and predictive capabilities of this model, we apply it in empirical case studies in Section 4.

4 Model Validation

Sales data for 253 products obtained from two FMCG companies in Australia are considered to investigate the validity and practical relevance of our proposed methodology. Both companies are major players in the food and beverage industry who supply most retailers nationally. We refer to these companies as Company A and Company B for anonymity reasons. In both cases, the supply chain consists of a single supplier (i.e., the case company) satisfying the demands of multiple retailers across the country. Time series data was provided by both companies in a weekly format including the actual sales, baseline statistical forecasts, final forecasts, and promotional mechanics. The datasets were amalgamated to perform our statistical testing. Results are reported on the entire large dataset so that there are enough observations to ensure generalizability.

The data has some notable characteristics. Both companies produce baseline statistical forecasts using an exponential smoothing model – which has shown to be a common industry practice (Hyndman et al., 2008). These baseline forecasts are regularly adjusted by a panel of experienced forecasters to consider the impact of promotions and other contextual information in the final forecast. We refer to the forecast generated by the companies as "Case forecast". The forecasting teams within each company consist of several experts, with their work experience varying from 2 to 20 years. There is, however, no specific procedure in place for identifying and modeling different types of promotions or structuring contextual information of any type. The DUS algorithm and FSE model can be utilized to tackle this matter.

Although the case companies produce their forecasts in a similar fashion, the shelf life (i.e., a function of product perishability) and frequency of promotions vary from one product to another. For example, Company A's products are less frequently promoted but the impact on sales is larger (i.e., promotional sales can exceed 60 times larger than non-promotional sales). Company B's products are more frequently promoted but the impact on sales is smaller compared to Company A. The natural logarithm of sales data was used in our model to stabilize the variance. Testing our new forecasting methodology on numerous products with differing demand patterns demonstrates its applicability, rigor and efficacy.

The DUS algorithm was first run to identify different demand states. ANOVA revealed that 'promotion type' and 'display type' are the most significant factors during promotions for Company A. This verified the opinion of forecasting experts at Company A who stated that these were some of the most influential factors affecting promotional sales. 'Advertising type' did not significantly change the promotional sales and it was, therefore, not included as a promotional variable in this model[4]. Promotion type relates to the depth of the discount offered and can take on one of the two possible levels: 'major' and 'minor'. Major and minor promotions are typically associated with discounts of approximately 50% and 30% off regular price, respectively. The 'display type' factor can also take one of the four possible levels including: 'entrance', 'Front Gondola End (FGE)', 'other gondola', or 'fixture'. Display type relates to the location in the store where products are displayed. Entrance display

[4] The advertisement type for Company A was not significant because both major and minor promotions for the investigated products were advertised in the retailer's local weekly catalogue. In addition to the local advertisement, major promotions were advertised by default as part of the retailer's national campaign and they could not be separated into individual factors. There were no other advertisement types utilized by Company A for the products investigated.

types are special out of aisle product displays positioned near the entrance of retail stores. Entrance displays are the most expensive in-store location for product placement but have the most impact on customer purchasing behavior. The FGE displays are at the end of aisles located nearest the front of the store, whereas other gondola displays may be at the end of an aisle but located near the back of the store. Fixture displays are in the aisle of the store that the product is normally located, where there is no additional shelf space for the promotional displays.

Table 1 illustrates a sample of the possible states for a product at Company A. Although, theoretically, there are eight possible combinations of the levels of the two promotional types, based on available data, the DUS algorithm prescribes five uplift states. The empty cells in Table 1 correspond to infeasible combinations in practice. For instance, the combination of 'minor' promotion type and 'entrance' in display type was not an allowed (feasible) combination by the company. That is, only products that were being offered on 'major' promotion were displayed at the store entrance. The DUS algorithm was implemented on other products and corresponding states were determined accordingly. Note that the promotional states and magnitude of sales uplift may slightly differ for various products.

Table 1: Promotional states identified by the DUS algorithm (average sales uplift in parentheses)

Promotion Type	Display Type			
	Entrance	FGE	Other Gondola	Fixture
Major	State 1 (19816)	State 2 (14833)	-	-
Minor	-	State 3 (5091)	State 4 (3466)	State 5 (4121)

Promotion type' and 'advertisement type' are the only two promotional mechanics found to be significant variables in our model for Company B (confirmed by forecasting experts and ANOVA). The 'promotion type' factor can take on one of the two possible levels including 'single buy' and 'multiple buy'. Promotion type relates to whether a promotional discount is offered for a single product or multiple product purchases (e.g., buy one for $25, two for $40, or three for $50). 'Advertisement type' relates to how/where the promotion is advertised, for which the variable can take on one of the three possible levels: 'catalogue', 'minor catalogue' or 'in-store'. Both promotional mechanics prove to be statistically significant at the 5% level. Although there are six possible combinations of the levels of the two promotion types, based on available data, the DUS algorithm prescribes five demand uplift states, as shown in Table 2. According to the DUS algorithm, two different combinations including the 'single buy' promotion type with the 'in-store' advertisement type as well as the 'multiple buy' promotion type with the 'in-store' advertisement type are grouped in State 4.

Table 2: Promotional states identified by the DUS algorithm (average sales uplift values in parentheses)

Promotion Type	Advertisement Type		
	Catalogue	In-Store	Minor Catalogue
Single buy	State 1 (311)	State 4 (16)	State 2 (213.3)
Multiple buy	State 3 (160.48)	State 5 (15.8)	State 6 (7.3)

Once the states were identified using the DUS algorithm, the FSE model was utilized to generate forecasts. The first 80 weeks of historical data was used as a training-set, and the last 20 weeks as a test-set to evaluate the performance of the FSE model. More precisely, we fit the FSE model to only include the first 80 weeks of sales data for both Companies A and B products and then use this fitted model to forecast demand for the next 20 time periods. The KPSS test was implemented on the baseline time series to test the stationary of time series at the significance level of 5%. An ARIMA model of order p and q that yielded the lowest AICc, was chosen as the best fit. The orders are automatically selected with using the Hyndman and Khandakar (2007) algorithm. Thus, we set the order parameters p and q in the FSE model. The range of p and q vary between one and three. The model was coded and implemented in the R 3.6.1 programming language where parameters are estimated, and forecasts are generated. To demonstrate the validity of the fitted FSE model, diagnostic tests and residual analyses were performed (i.e., independent and homoscedastic residuals, normality, Ljung-Box test to check the fitness of models at a 5% level of significance).

The accuracy of forecasts is analyzed in the following subsection using different forecast accuracy metrics as well as benchmarking the forecasts produced by the FSE model against some of the sophisticated demand forecasting techniques.

4.1 Model accuracy and benchmarking

Different criteria have been used by researchers and practitioners to gauge the effectiveness of a forecasting model. It is of utmost importance to select appropriate error measures when evaluating forecast accuracy (Davydenko & Fildes, 2013). Since the scale of sales dramatically changes and differs for the investigated products, we explore the performance of models with scale independent metrics. To ensure consistency and comparability with previous studies (e.g., Baecke, De Baets & Vanderheyden, 2017; Fildes et al., 2009; Hyndman & Koehler, 2006; Kourentzes & Petropoulos, 2016), we utilize Mean Absolute Scaled Error (MASE) as the primary accuracy measure in our analyses as it is scale independent. The MASE is computed using Equation 2,

$$MASE = \frac{1}{n}\sum_{t=1}^{n}\frac{|f_t - x_t|}{\sum_{t=2}^{m}\frac{|x_t - x_{t-1}|}{m-1}}, \tag{2}$$

where f_t is the forecast value at time t, x_t is the actual sales at time t, and n is the length of out-of-sample size. MASE is a metric that assesses the accuracy of forecasts against a naïve forecasting model. The denominator of Equation 2 is a simple naive model where m is the length of within sample size.

The forecasts obtained from the FSE model are also benchmarked against some of the common demand forecasting techniques to gauge its effectiveness. The well-known ARIMA model is chosen as a widely accepted standard benchmark for product demand forecasting (Hyndman & Athanasopoulos, 2014). We also benchmark against ARIMAX as an extension of its univariate form, ARIMA, which accounts for promotional information. ARIMAX is implemented automatically using the 'TSA' package in R (Chan et al., 2018).

Support Vector Regression (SVR) and Regression Trees (RT) are implemented as two common and successful machine learning models that have been employed in numerous studies to forecast promotional sales (e.g., Abolghasemi et al., 2020; Ali et al., 2009). SVR employs a kernel function to transfer low dimensional data to high dimensional data where the kernel function can take on several different forms. Different kernel functions were trialed including poly, RBF, and a linear function. A grid search was also used to choose the best combination of the kernel function, kernel coefficient (γ) and cost function (C). A sequence of intervals of width 0.1 (ranging from 0 to 1) were ran for γ, and on a sequence of intervals of width 1 (ranging from 0 to 100) for C. The model was fit using the 'SVM' function from the 'e1071' package in R (Dimitriadou et al., 2006) and the model that resulted in the highest accuracy was selected. An RT is a decision tree that splits the dataset into smaller groups called branches and fits a linear regression to each group. RT models are a powerful class of Classification and Regression Tree (CART) that have been successfully used for demand forecasting (Breiman, 2017). We fit RT using 'RT' function with 'ANOVA' splitting in the 'Rpart' package in R (Therneau et al., 2015).

4.2 Empirical results

The summary results of forecasting models and the significance of differences are reported in Table 3 and Table 4, respectively. Notably, the average accuracy of models is close to one another with FSE model outperforming other models, on average. The FSE model outperforms the standard ARIMA model. This is not surprising as ARIMA does not account for promotional information while our model factors in promotional information as external regressors. The results obtained from our model outperform those obtained from ARIMAX and the commonly practiced judgmental techniques employed by the case companies. The results for mean of MASE is different from median of MASE. While the Case forecasts are more accurate than SVR and RT, on average, the SVR and RT models have higher accuracy in terms of median MASE. However, the FSE model still outperforms others in terms of median of MASE. This indicates that DUS algorithm has effectively defined the states of systematic events utilized in the FSE model and avoids generating poor forecasts.

Table 3: Forecast accuracy in the test-set period

	FSE	ARIMAX	Case	SVR	RT	ARIMA
MASE (Mean)	0.618	0.643	0.618	0.698	0.622	0.691
MASE (Median)	0.458	0.486	0.485	0.477	0.465	0.578

Table 4: Comparing the performance of forecasting models (p-values of pairwise t-tests)

	ARIMAX	Case	SVR	RT	ARIMA
FSE	0.0467*	0.014*	0.238	0.938	0.015*
ARIMAX		0.008*	0.446	0.776	0.008*
Case			0.001*	0.013*	0.063**
SVR				0.271	0.001*

RT	0.014*

*Significant at α = 0.05
**Significant at α = 0.1

From Table 4, there is a statistically significant difference between the results of the FSE model and those obtained from ARIMAX, the case companies, and ARIMA. However, there is no significant difference between the FSE and machine learning methods (SVR and RT). Although, the FSE method is simpler to apply and yet produces more accurate forecasts, on average, while it also better handles extreme observations. This is also evident in Figure 1.

Figure 1 shows the scatter plot of forecast accuracy for different models with respect to normalized sales. We normalized sales by finding the average for each series; enabling the performance of different models to be compared across series with different scales. From Figure 1, while the FSE model produces competitive forecasts, on average, it avoids generating poor forecasts. The fact that the FSE model shows the most robust performance is commendable since one important aspect of forecasting is to avoid wild forecasts that substantially miss the actual values (Shaub, 2020). It can be seen that the ARIMA, SVR and RT models have significantly greater risk of poor performance than other models despite producing relatively accurate forecasts in many occasions. We further investigated the series that have performed poorly (i.e., the series with MASE mean greater than 2). There are 31 unique series in total that corresponds to 31 products. One common feature among all these series is that they all have at least one major promotional period in the test set. We observe that the forecasting model misbehaves in the promotional periods. This causes a large error on average, despite the relative accuracy of the model in non-promotional periods. Note that the scale of MASE differs for different models. For example, the forecasts that are generated in the case studies and the FSE model has a smaller scale of MASE. Both the case company techniques and FSE models use judgment to forecast sales. While the former uses judgment *subjectively*, the latter uses the DUS algorithm to *objectively* consider promotional information in the model. This is a considerable achievement and a significant step towards the objective use of human judgment in forecasting (Goodwin, 2020).

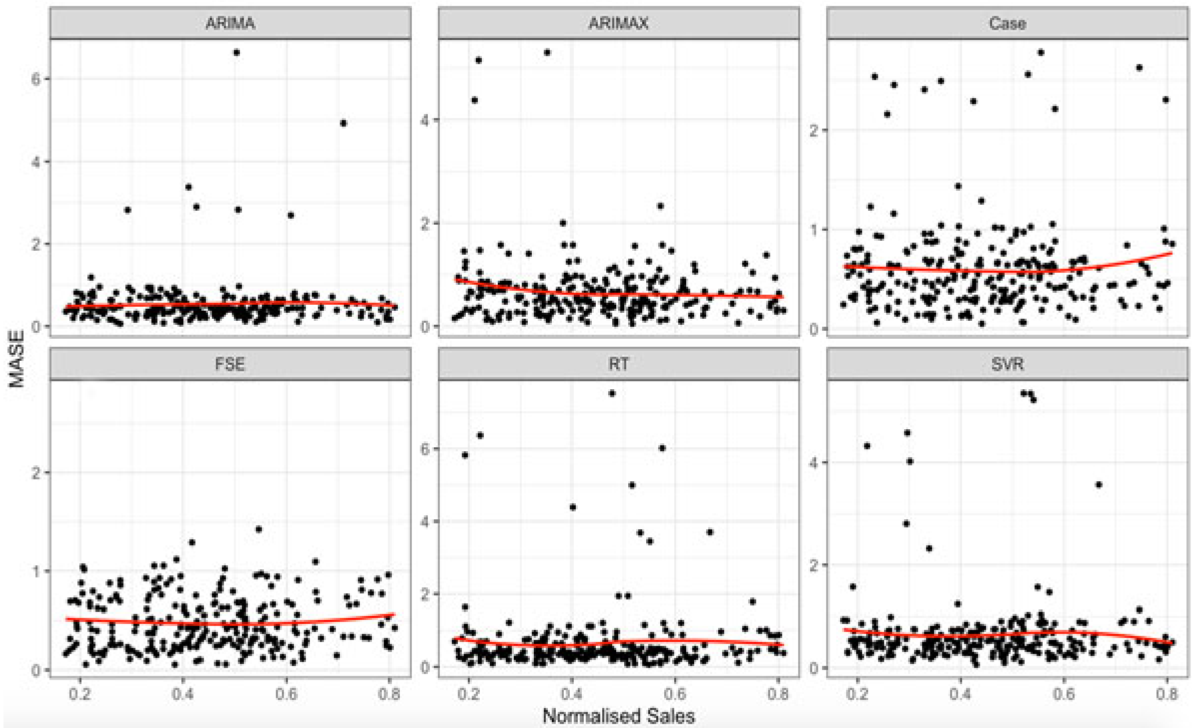

Figure 1: Forecast accuracy (MASE) of different models

4.3 Discussion

Section 4.2 showed the robustness of the FSE model featuring consistency in generating accurate forecasts while avoiding forecasts with large errors – a competitive advantage over those benchmarked against. We now turn our attention to determine the conditions upon which the FSE model could outperform the others. Specifically, we investigate the promotion impact on the forecasting accuracy of models because the FSE model is motivated by modeling systematic events (e.g., promotional information), which is also the primary focus of this paper. Since the nature of series and scales of sales are different during promotional periods than non-promotional periods, the promotional observations are separated from non-promotional ones and the results are analyzed individually. There is a total of 4831 promotional observations in the training set and there are 966 promotional observations over the test set.

Table 5 shows the forecast accuracy of models in the presence (absence) of promotions and Table 6 shows the significance of difference between forecasting models after applying pairwise t-tests. As shown in Table 5, all models have higher accuracy and perform better in the absence of promotions.

The results are however different during the promotional periods. The FSE model outperforms other models during the promotions, while it is ranked third for non-promotional periods. On the other hand, the ARIMA model has the highest accuracy during non-promotional periods despite the poor performance during promotional periods. This indicates how significantly promotions can impact the performance of forecasting models, and more importantly how effectively the FSE model can outperform the other methods in forecasting the promotional sales.

Table 5: Forecast accuracy (mean of MASE) in the test-set period (rank in parenthesis)

	FSE	ARIMAX	Case	SVR	RT	ARIMA
Promo	1.457 (I)	1.650 (II)	1.750 (IV)	1.751 (V)	1.686 (III)	1.784 (VI)
Non-Promo	0.245 (II)	0.274 (IV)	0.274 (IV)	0.276 (VI)	0.273 (III)	0.235 (I)

Table 6: Comparing the performance of forecasting models in the presence of promotions (p-values of pairwise t-tests; promotional periods in upper-right half and non-promotional periods in lower-left half of table)

	FSE	ARIMAX	Case	SVR	RT	ARIMA
FSE		0.009*	0.080**	0.989	0.568	0.013*
ARIMAX	0.039*		0.035*	0.393	0.741	0.010*
Case	0.004*	0.079**		0.082**	0.198	0.043*
SVR	0.771	0.476	0.004*		0.563	0.013*
RT	0.367	0.611	0.027*	0.230		0.036*
ARIMA	0.190	0.018*	0.001*	0.004*	0.053*	

*Significant at $\alpha = 0.05$
**Significant at $\alpha = 0.1$

From Table 5, we can see that the performance of different models changes during promotional and non-promotional periods. Some methods may not be able to capture the underlying demand for spikes or troughs, and consequently, may overfit or misbehave as they encounter these observations (Ord, 2020). However, our proposed method benefits from regime-switching approach and uses promotional information objectively to estimate the promotional sales and effectively switches to another model (ARIMA) to forecast the baseline sales. This approach avoids any forecast error to be carried forward from the previous observations. The significance of this approach is more pronounced during promotions when forecast misbehavior would be more costly to the supply chain.

The gray cells on the lower half of Table 6 show the p-values of paired t-tests to test if there is any significant difference between each pair of models during non-promotional periods. The upper half of the table shows if there is any significant difference between each pair of models during promotional periods. The results are dissimilar for promotional and non-promotional periods. During promotional periods, the ARIMA model produces significantly different (worse) forecasts from other models. The FSE model generates significantly more accurate forecasts than the ARIMAX, the case companies, and the ARIMA models. The case companies model has a significantly different performance to all other model except ARIMAX. However, there is no significant difference between the forecasts generated by SVR and RT and those generated by the FSE model. The similarity of RT and FSE is interesting since the DUS algorithm used in the FSE model mimics a similar approach to RT that tries to partition the dataset to similar subsets. During non-promotional periods, the FSE generates significantly different forecasts compared to the others, but again there is no statistically significant difference between FSE forecasts and the forecasts of ARIMA, SVR and RT models. Despite the machine learning methods appearing to effectively capture the effects of sales promotions, these methods can be highly complicated in real world applications; hence not always the preferred methods in industry.

Our aim in this paper was to provide a robust, yet simple and practical, method to effectively model systematic events and produce reliable forecasts. The approach allows experts' judgment to be objectively structured (using the DUS algorithm) in a regime-switching approach to forecast time series that are impacted by systematic events (e.g., sales during promotional and non-promotional periods). We developed a pragmatic approach (DUS algorithm and FSE model) that can incorporate promotional information objectively into a forecasting model whose advantage is not just to generate reliable, competitive forecasts, but to also avoid wild forecasts with large errors caused by the series entering different states. The FSE model reduces the forecast error in promotional and non-promotional periods by effectively identifying various demand states and using appropriate model that takes into account the relevant information to forecast sales.

5 Conclusions

In this paper, we propose a time series regression model that structures and embeds systematic contextual information that would otherwise be incorporated with unaided human judgment. Structuring is achieved using an approach, called DUS algorithm, that systematically defines demand uplift states for combinations of levels of factors that contribute to demand uplifts. The factors to consider in DSU algorithm are provided by expert forecasters allowing the model to also benefit from expert knowledge. Once the demand uplift states are identified, they are incorporated into a forecasting model, called FSE, which considers the impact of systematic events to forecast the future demand.

The proposed model and methodology were applied to prepare forecasts for 253 products/time series. We find that the systematic event structuring and forecasting approach can remarkably improve accuracy when compared to the judgmentally made forecasts by the company forecasters (i.e., experienced forecasters making judgmental adjustments to baseline statistical forecasts). Forecast accuracy improvements were demonstrated from the companies' dataset using MASE and benchmarking techniques. We consider this as the key practical contribution of this model since the improved forecast accuracy brings about substantial cost savings according to our industry partners. Indeed, for FMCG companies who forecast for thousands of products on a routine basis, more accurate forecasts – through minimizing judgmental adjustments to incorporate systematic events – can save substantial time and money.

The FSE model is simple and practical in generating competitive and robust forecasts. This model is also shown to effectively capture various demand states which is essential in avoiding extremely poor forecasts (i.e., wild forecasts with large errors). One advantage of the FSE model is the ease with which it allows translating systematic information into demand states. We show in this paper that this can be an effective approach to account for the impact of systematic events such as sales promotions. The difficulty, however, is that the DUS algorithm used to identify demand uplift states is not automated and hence requires substantial time investment when forecasting for a large number of products. We see automating this process – for example, using feedback systems and machine learning approaches – as one important direction for future research. The positive side is that the DUS algorithm that informs the forecasting model does not need to be executed over and over in every forecasting period. Once the demand uplift states are identified, the FSE model can just rerun using the fixed states until new events with different characteristics appear which may then the DUS algorithm to be updated. Another limitation of the proposed approach is the need to have access to sufficient historical data to identify demand uplift states. And for this reason, the FSE model may not be well suited for new products for which the historical sales data is unavailable. This seems to be a common limitation of most statistical models which reply on historical data.

Despite the ease of use, the FSE model can improve forecast accuracy through mitigating the use of unstructured information which seems to be an undeniable part of judgmental forecast adjustments in practice. Part of the uncertainties in demand forecasting roots in how different forecasters evaluate the potential effects of contextual information. For example, different forecasters may have different, potentially contrasting, opinions about the impacts of various promotions. Variations could be caused by psychological reasons or could be due to different experiences, market and supply chain knowledge, access to information, or targets to achieve (Oliva & Watson, 2009). Our model helps overcome such issues as it allows a forecaster to objectively capture the corresponding effects of different quantifiable systematic events such as promotions.

The model and approach developed in this paper is the first attempt to apply concept of regime-switching to define demand states to capture the effects of systematic events. The resulting forecast could be obviously further adjusted by the forecasters to incorporate less quantifiable contextual information and the impact of events that cannot be systematically formulated. Our study thus contributes to the call for further research in structuring the use of human judgment in forecasting (Green & Armstrong, 2007), particularly for products that are prone to sporadic perturbations (De Baets & Harvey, 2018). In the future, the FSE model could also be applied and tested on data at a more disaggregate level (e.g., store, town and/or district level).

Our research also provides insights for innovative FSS design, especially the need for structured support to assist with filtering and integrating information (Fildes, Goodwin & Önkal, 2018). Future research may investigate how this model could be embedded in an FSS so that forecasts for a large number of products are produced semi-automatically. By doing so, judgmental adjustments to forecasts can be more systematic and hassle-free (the use of an FSS can help reduce the cognitive burden on forecasters), and the chance of accounting for the same information twice in demand planning and S&OP would be reduced.

Acknowledgement

This study was funded by the Australian Research Council (Grant ID: IC140100032). The authors are grateful to Prof Enno Siemsen from Wisconsin School of Business, Prof Marcus O'Connor from The University of Sydney and Prof Elliot Bendoly from The Ohio State University for their invaluable feedback and constructive comments throughout this research.

References

Abolghasemi, M, Beh, E, Tarr, G & Gerlach, R 2020, 'Demand forecasting in supply chain: The impact of demand volatility in the presence of promotion', *Computers & Industrial Engineering*, pp. 106380.

Adya, M & Lusk, EJ 2016, 'Development and validation of a rule-based time series complexity scoring technique to support design of adaptive forecasting DSS', *Decision Support Systems,* vol. 83, pp. 70-82.

Ali, ÖG, Sayın, S, Van Woensel, T & Fransoo, J 2009, 'SKU demand forecasting in the presence of promotions', *Expert Systems with Applications,* vol. 36, no. 10, pp. 12340-12348.

Alvarado-Valencia, J, Barrero, LH, Önkal, D & Dennerlein, JT 2017, 'Expertise, credibility of system forecasts and integration methods in judgmental demand forecasting', *International Journal of Forecasting,* vol. 33, no. 1, pp. 298-313.

Armstrong, JS 1986, 'The ombudsman: research on forecasting: A Quarter-Century Review, 1960–1984', *Interfaces,* vol. 16, no. 1, pp. 89-109.

Baecke, P, De Baets, S & Vanderheyden, K 2017, 'Investigating the added value of integrating human judgement into statistical demand forecasting systems', *International Journal of Production Economics,* vol. 191, pp. 85-96.

Blattberg, RC & Hoch, SJ 1990, 'Database models and managerial intuition: 50% model+ 50% manager', *Management Science,* vol. 36, no. 8, pp. 887-899.

Blattberg, RC & Neslin, SA 1989, 'Sales promotion: The long and the short of it', *Marketing letters,* vol. 1, no. 1, pp. 81-97.

Breiman, L 2017, *Classification and regression trees*, Routledge.

Chan, K-S, Ripley, B, Chan, MK-S & Chan, S 2018, 'TSA: Time Series Analysis. R package version 1.2'.

Cooper, LG, Baron, P, Levy, W, Swisher, M & Gogos, P 1999, 'PromoCast™: A new forecasting method for promotion planning', *Marketing Science,* vol. 18, no. 3, pp. 301-316.

Davydenko, A & Fildes, R 2013, 'Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts', *International Journal of Forecasting,* vol. 29, no. 3, pp. 510-522.

De Baets, S & Harvey, N 2018, 'Forecasting from time series subject to sporadic perturbations: Effectiveness of different types of forecasting support', *International Journal of Forecasting,* vol. 34, no. 2, pp. 163-180.

Dimitriadou, E, Hornik, K, Leisch, F, Meyer, D, Weingessel, A & Leisch, MF 2006, 'The e1071 package', *Misc Functions of Department of Statistics (e1071), TU Wien*.

Divakar, S, Ratchford, BT & Shankar, V 2005, 'CHAN4CAST: A Multichannel, Multiregion Sales Forecasting Model and Decision Support System for Consumer Packaged Goods', *Marketing Science,* vol. 24, no. 3, pp. 334-350.

Edmundson, B, Lawrence, M & O'connor, M 1988, 'The use of non-time series information in sales forecasting: A case study', *Journal of Forecasting,* vol. 7, no. 3, pp. 201-211.

Fildes, R 1992, 'The evaluation of extrapolative forecasting methods', *International Journal of Forecasting,* vol. 8, no. 1, pp. 81-98.

Fildes, R & Goodwin, P 2007, 'Against your better judgment? How organizations can improve their use of management judgment in forecasting', *Interfaces,* vol. 37, no. 6, pp. 570-576.

Fildes, R, Goodwin, P & Lawrence, M 2006, 'The design features of forecasting support systems and their effectiveness', *Decision Support Systems,* vol. 42, no. 1, pp. 351-361.

Fildes, R, Goodwin, P, Lawrence, M & Nikolopoulos, K 2009, 'Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning', *International journal of forecasting,* vol. 25, no. 1, pp. 3-23.

Fildes, R, Goodwin, P & Önkal, D 2018, 'Use and misuse of information in supply chain forecasting of promotion effects', *International Journal of Forecasting*.

Fildes, R, Nikolopoulos, K, Crone, SF & Syntetos, A 2008, 'Forecasting and operational research: a review', *Journal of the Operational Research Society,* vol. 59, no. 9, pp. 1150-1172.

Fischer, I & Harvey, N 1999, 'Combining forecasts: What information do judges need to outperform the simple average?', *International journal of forecasting,* vol. 15, no. 3, pp. 227-246.

Franses, PH 2008, 'Merging models and experts', *International Journal of Forecasting,* vol. 24, no. 1, pp. 31-33.

Franses, PH & Legerstee, R 2013, 'Do statistical forecasting models for SKU-level data benefit from including past expert knowledge?', *International Journal of Forecasting,* vol. 29, no. 1, pp. 80-87.

Gardner, ES 1985, 'Exponential smoothing: The state of the art', *Journal of forecasting,* vol. 4, no. 1, pp. 1-28.

Giordani, P & Kohn, R 2008, 'Efficient Bayesian inference for multiple change-point and mixture innovation models', *Journal of Business & Economic Statistics,* vol. 26, no. 1, pp. 66-77.

Giordani, P, Kohn, R & Van Dijk, D 2007, 'A unified approach to nonlinearity, structural change, and outliers', *Journal of Econometrics,* vol. 137, no. 1, pp. 112-133.

Goodwin, P 2002, 'Integrating management judgment and statistical methods to improve short-term forecasts', *Omega,* vol. 30, no. 2, pp. 127-135.

Goodwin, P 2020, 'Performance measurement in the M4 Competition: Possible future research', *International Journal of Forecasting,* vol. 36, no. 1, pp. 189-190.

Goodwin, P, Fildes, R, Lawrence, M & Stephens, G 2011, 'Restrictiveness and guidance in support systems', *Omega,* vol. 39, no. 3, pp. 242-253.

Green, KC & Armstrong, JS 2007, 'Structured analogies for forecasting', *International Journal of Forecasting,* vol. 23, no. 3, pp. 365-376.

Green, KC & Armstrong, JS 2015, 'Simple versus complex forecasting: The evidence', *Journal of Business Research,* vol. 68, no. 8, pp. 1678-1685.

Hamilton, JD 1989, 'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica: Journal of the Econometric Society*, pp. 357-384.

Hamilton, JD 1990, 'Analysis of time series subject to changes in regime', *Journal of econometrics,* vol. 45, no. 1-2, pp. 39-70.

Hogarth, RM & Makridakis, S 1981, 'Forecasting and planning: An evaluation', *Management science,* vol. 27, no. 2, pp. 115-138.

Huang, T, Fildes, R & Soopramanien, D 2014, 'The value of competitive information in forecasting FMCG retail product sales and the variable selection problem', *European Journal of Operational Research,* vol. 237, no. 2, pp. 738-748.

Hughes, M 2001, 'Forecasting practice: organisational issues', *Journal of the Operational Research Society,* vol. 52, no. 2, pp. 143-149.

Hyndman, RJ & Athanasopoulos, G 2014, *Forecasting: principles and practice*, OTexts.

Hyndman, RJ & Khandakar, Y 2007, *Automatic time series for forecasting: the forecast package for R*, Monash University, Department of Econometrics and Business Statistics.

Hyndman, RJ & Koehler, AB 2006, 'Another look at measures of forecast accuracy', *International journal of forecasting,* vol. 22, no. 4, pp. 679-688.

Hyndman, RJ, Koehler, AB, Ord, JK & Snyder, RD 2008, *Forecasting with exponential smoothing: the state space approach*, Springer Science & Business Media.

Kim, C-J, Piger, J & Startz, R 2008, 'Estimation of Markov regime-switching regression models with endogenous switching', *Journal of Econometrics,* vol. 143, no. 2, pp. 263-273.

Kourentzes, N & Petropoulos, F 2016, 'Forecasting with multivariate temporal aggregation: The case of promotional modelling', *International Journal of Production Economics,* vol. 181, pp. 145-153.

Kremer, M, Siemsen, E & Thomas, DJ 2015, 'The sum and its parts: Judgmental hierarchical forecasting', *Management Science,* vol. 62, no. 9, pp. 2745-2764.

Lawrence, M, Goodwin, P, O'connor, M & Önkal, D 2006, 'Judgmental forecasting: A review of progress over the last 25 years', *International Journal of Forecasting,* vol. 22, no. 3, pp. 493-518.

Lawrence, M, O'connor, M & Edmundson, B 2000, 'A field study of sales forecasting accuracy and processes', *European Journal of Operational Research,* vol. 122, no. 1, pp. 151-160.

Makridakis, S, Andersen, A, Carbone, R, Fildes, R, Hibon, M, Lewandowski, R, Newton, J, Parzen, E & Winkler, R 1982, 'The accuracy of extrapolation (time series) methods: Results of a forecasting competition', *Journal of forecasting,* vol. 1, no. 2, pp. 111-153.

Makridakis, S, Chatfield, C, Hibon, M, Lawrence, M, Mills, T, Ord, K & Simmons, LF 1993, 'The M2-competition: A real-time judgmentally based forecasting study', *International Journal of Forecasting,* vol. 9, no. 1, pp. 5-22.

Makridakis, S & Hibon, M 2000, 'The M3-Competition: results, conclusions and implications', *International journal of forecasting,* vol. 16, no. 4, pp. 451-476.

Marmier, F & Cheikhrouhou, N 2010, 'Structuring and integrating human knowledge in demand forecasting: a judgemental adjustment approach', *Production Planning and Control,* vol. 21, no. 4, pp. 399-412.

Moritz, B, Siemsen, E & Kremer, M 2014, 'Judgmental forecasting: Cognitive reflection and decision speed', *Production and Operations Management,* vol. 23, no. 7, pp. 1146-1160.

Nikolopoulos, K, Litsa, A, Petropoulos, F, Bougioukos, V & Khammash, M 2015, 'Relative performance of methods for forecasting special events', *Journal of Business Research,* vol. 68, no. 8, pp. 1785-1791.

Nikolopoulos, K, Syntetos, AA, Boylan, JE, Petropoulos, F & Assimakopoulos, V 2011, 'An aggregate–disaggregate intermittent demand approach (ADIDA) to forecasting: an empirical proposition and analysis', *Journal of the Operational Research Society,* vol. 62, no. 3, pp. 544-554.

Oliva, R & Watson, N 2009, 'Managing functional biases in organizational forecasts: A case study of consensus forecasting in supply chain planning', *Production and operations Management,* vol. 18, no. 2, pp. 138-151.

Ord, K 2020, 'Data adjustments, overfitting and representativeness', *International Journal of Forecasting,* vol. 36, no. 1, pp. 195-196.

Ramanathan, U 2012, 'Supply chain collaboration for improved forecast accuracy of promotional sales', *International Journal of Operations & Production Management,* vol. 32, no. 6, pp. 676-695.

Ramanathan, U & Muyldermans, L 2010, 'Identifying demand factors for promotional planning and forecasting: A case of a soft drink company in the UK', *International journal of production economics,* vol. 128, no. 2, pp. 538-545.

Ramanathan, U & Muyldermans, L 2011, 'Identifying the underlying structure of demand during promotions: A structural equation modelling approach', *Expert systems with applications,* vol. 38, no. 5, pp. 5544-5552.

Sanders, NR 1992, 'Accuracy of judgmental forecasts: A comparison', *Omega,* vol. 20, no. 3, pp. 353-364.

Sanders, NR & Manrodt, KB 2003a, 'The efficacy of using judgmental versus quantitative forecasting methods in practice', *Omega,* vol. 31, no. 6, pp. 511-522.

Sanders, NR & Manrodt, KB 2003b, 'Forecasting software in practice: Use, satisfaction, and performance', *Interfaces,* vol. 33, no. 5, pp. 90-93.

Seifert, M, Siemsen, E, Hadida, AL & Eisingerich, AB 2015, 'Effective judgmental forecasting in the context of fashion products', *Journal of Operations Management,* vol. 36, pp. 33-45.

Shaub, D 2020, 'Fast and accurate yearly time series forecasting with forecast combinations', *International Journal of Forecasting,* vol. 36, no. 1, pp. 116-120.

Taylor, JW 2003, 'Exponential smoothing with a damped multiplicative trend', *International journal of Forecasting,* vol. 19, no. 4, pp. 715-725.

Therneau, T, Atkinson, B, Ripley, B & Ripley, MB 2015, 'Package 'rpart'', *Available online: cran. ma. ic. ac. uk/web/packages/rpart/rpart. pdf (accessed on 20 April 2016)*.

Trapero, JR, Kourentzes, N & Fildes, R 2015, 'On the identification of sales forecasting models in the presence of promotions', *Journal of the operational Research Society,* vol. 66, no. 2, pp. 299-307.

Trapero, JR, Pedregal, DJ, Fildes, R & Kourentzes, N 2013, 'Analysis of judgmental adjustments in the presence of promotions', *International Journal of Forecasting,* vol. 29, no. 2, pp. 234-243.

Van Heerde, HJ, Leeflang, PS & Wittink, DR 2002, 'How promotions work: SCAN* PRO-based evolutionary model building', *Schmalenbach Business Review,* vol. 54, no. 3, pp. 198-220.

Webby, R & O'connor, M 1996, 'Judgemental and statistical time series forecasting: a review of the literature', *International Journal of Forecasting,* vol. 12, no. 1, pp. 91-118.